Use of Genomic Biomarkers in a Regulatory Environment

Robert J. Temple, M.D.
Associate Director for Medical Policy
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

DIA: Genomic Biomarkers
October 7, 2005

Roles of Biomarkers

Biomarkers can define or explain a disease or assess its severity, examine a drug's ability to bind to and influence a receptor, encourage development of a drug that seems to move the marker in the right direction or stop development of a drug that does not.

Today, I want to talk about two particular roles

1. A biomarker that assesses disease severity or risk.
2. A biomarker that predicts response to treatment.

I'll touch on surrogate endpoints but biomarkers will be of great value even if they don't achieve that status because they will allow us to conduct more efficient trials in enriched populations.

2

Genomic vs Other Biomarkers

Interest today is in the new opportunities for product development that genomic biomarkers can bring us and there clearly is anxiety about the new technology involved.

But the uses of genomic biomarkers are not, I believe different conceptually from other kinds of biomarkers, as well as physical or clinical and historical differences that distinguish one person from another. That should, I think, be reassuring, although certainly details will matter.

3

Biomarkers: Definition

A biomarker is any measurable physiologic, pathologic, structural (micro or macro), or genetic property that can define or measure a physiologic or pathologic activity, or influence or predict a disease process, either by its presence (risk factors) or by its response to a treatment.

Biomarkers include:

- Basic properties of a cell (binding of a molecule, activity of K, Na, Ca channels)
- Concentration or activity/specificity of an enzyme (renin activity)
- Circulating molecules that predict disease presence or severity (hormone levels, creatinine lipids, CRP, HgAlc, bilirubin, placental growth factor, troponin, PSA)

4

Biomarker: Definition

- Anatomic abnormalities (coronary plaque, brain ventricle size, tumor size, MRI findings)
- Functional activity (glucose uptake, binding of antibody, echocardiographic Tc scan finding
- Genetic markers in whole body (risk of Alzheimer's Disease, breast or ovarian cancer) or tumor (define aggressiveness, likelihood of metastasis)
- Tumor markers (EGFR, HER-2, abnormal tyrosine kinase)
- Critical history (recent AMI, TIA)
- Clinical measurements (BP, QT interval, HR)

5

Individualization

Enrichment of studies, selecting patients for a trial who are more likely to allow demonstration of an effect, fits with growing appreciation that people who seem similar may differ in their risk and likelihood of response to treatment. This is not new idea. Doctors have always felt that tailoring treatments was part of the "art of medicine." To do this they use a mixture of

- trial and error
- instinct
- experience
- knowledge, a growing component

A problem is that clinical trials are not usually designed to look at individual responses; they look at group effects.

6

Individualization

Where do we get information about differences?

Epidemiology gives us information about risk differences among people for various outcomes and even about interactions of risk factors, and it can relate the outcome to degree of abnormality of BP, cholesterol, etc.

Differences in response to treatment can sometimes be found in analyses of clinical trials showing how response relates to baseline characteristics. But these retrospective "subset analyses" have generally been treated with skepticism (except that we require demographic analyses in NDAs) and as "exploratory."

7

Subset Skepticism

Yusuf, Wittes, Probstfield, Tyroler [JAMA, 1991] described many risks of these analyses, despite recognizing desirability of knowing response differences:

- Multiple comparisons and increased alpha error; appropriate correction leaves very low power.
- Usually not prospective; post-facto analyses all look plausible but can be biased.

There are some famous errors.

8

Yusuf (cont'd)

Prudent to rely more on overall effect to decide on effect in a subgroup than on actual observation in the subgroup

Famous illustrations

1. GISSI - SK reduced mortality 20% but all effect in anterior MI. Subsequent studies and overview showed effects at all AMI sites

2. ISIS-2 - Aspirin beneficial overall and for persons under all zodiacal signs except Libra and Gemini, where it was harmful (Peto)

9

A Change in the Air

Beyond skepticism about subset analyses, there has been a philosophical view that even if subsets differ somewhat, the direction of effect will be similar and that public health need was to know how to treat everyone. The large simple trial reflects this. So – do LSTs, do not try too much to target treatment. View still seen in

- Polypill idea
- OTC statins

This may well be correct for some major interventions (e.g., BP) but I sense a change

10

A Change in the Air

Increasingly there is recognition, first, that there can be differences between people that can affect response importantly and that trying to identify those differences and target treatment to the people most likely to benefit may be desirable. Certainly, for drugs with significant toxicity (and perhaps significant cost) there is interest in discovering who really benefits.

Second, there is recognition that people differ in their risk of an event and that who really needs treatment can depend on this risk.

And discoveries of genetic and proteomic differences are advancing these changes.

11

Subset Analyses are the Norm (Cautiously)

Despite awareness of their risks (multiplicity), subset analyses (preferably planned in the protocol) are now routine in journal reports of successful outcome studies (you can't save a failed study by finding a successful subset) as so-called "forest plots," a vertical display of hazard ratios and CI's for various population subsets. They are also appearing in labeling (Tarceva, Toprol XL, carvedilol, many others). There has been particular interest in subsets defined by:

- Demographics
- Disease severity
- County/region
- Concomitant treatment
- Concomitant illness

12

Subset Analyses are the Norm (Cautiously)

They so far have not commonly displayed genomic/proteomic subsets but this is clearly coming (Tarceva).

The fact that people differ in their risk and response provides both an opportunity to target therapy better and also to utilize more efficient study designs in enriched populations.

13

Enrichment

Enrichment is prospective use of any patient characteristic – demographic, pathophysiologic, historical, or genetic, and others – to select patients for study to obtain a study population in which detection of a drug effect is more likely.

This occurs to a degree in virtually every trial and is intended to increase study power by:

- Decreasing heterogeneity
- Finding a population with many outcome events, i.e., high risk patients
- Identifying a population capable of responding to the treatment

14

Reducing Heterogeneity

These approaches are virtually universal

- Find (prospectively) likely compliers
- Choose people who will not drop out
- Eliminate placebo-responders in a lead-in period
- Eliminate people who give inconsistent treadmill results in heart failure or angina trials
- Eliminate people with diseases likely to lead to early death
- Eliminate people on drugs with the same effect as test drug

In general, these enrichments do not raise questions of generalizability.

15

Selection of High Risk Patients (more likely to have events)

Although the information distinguishing individuals is growing exponentially, we've had such information before

- Epidemiologic risk factors
  - Cholesterol, blood pressure levels
  - Diabetes
  - Prior events (AMI, stroke, PVD)
  - Family history
  - Gender, race, age
- Individual measurement/history
  - Previous breast cancer
  - Tumor histology
  - Arteriogram, echocardiogram, exercise testing
  - Evidence of MBD as predictor of Alzheimer's Disease

16

Selection of High Risk Patients

a. Oncology

- Tamoxifen prevented contralateral breast tumors in adjuvant setting (very high risk); it was then studied in people with more general high risk. This was needed a) to have enough endpoints to detect a possible effect and b) because of concern about toxicity. It was labeled for the group studied, with access to Gail Model calculator to assess risk. There was no reason in this case to expect larger % effect in the people selected, but more events would be prevented.

17

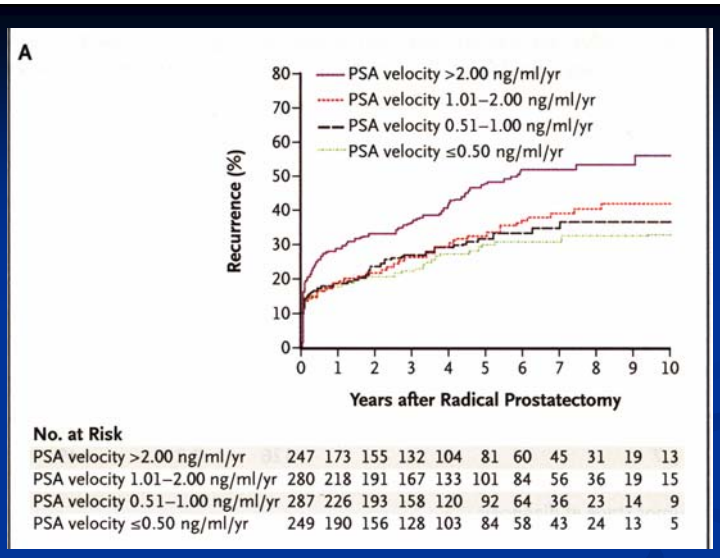
Selection of High Risk Patients

a. Oncology

- Potential selection method for frequent endpoints:

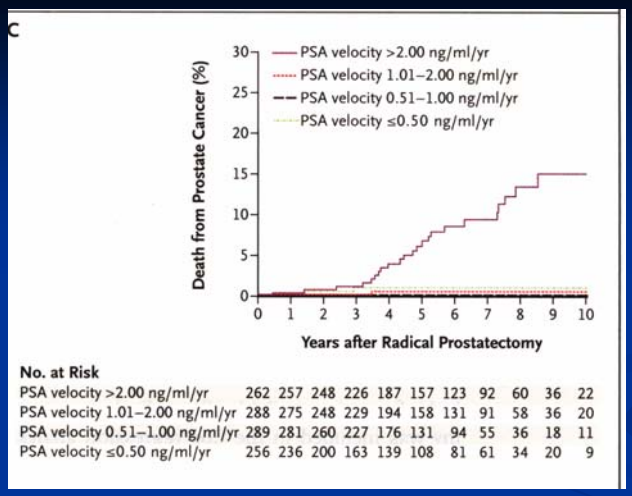
  D'Amico showed [NEJM 2004; 351:125-135] that in men with localized prostate Ca, following radical prostatectomy, PSA "velocity" (PSA increase > 2 ng/ml during prior year) identified virtually all patients who would die of prostate Ca over a 10 year period.

18



Kaplan-Meier Estimates of Disease Recurrence (Panel A) after Radical Prostatectomy, According to the Quartile of PSA Velocity during the Year before Diagnosis

19



Kaplan-Meier Estimates of the Cumulative Incidence of Death from Prostate Cancer (Panel C) after Radical Prostatectomy, According to the Quartile of PSA Velocity during the Year before Diagnosis

20

Selection of High Risk Patients

So, who would you put in your test of interventions at the time of radical prostatectomy to improve survival?

And could this have implications for treatment decisions?

21

Selection of High Risk Patients

a. Oncology

Oncologic predictors, both descriptive (where we don't understand the mechanism) and pathophysiologic are coming in droves

- Microarrays of SNP's predict likelihood of distant breast cancer metastases after surgery better than LN status, histology, tumor size, etc. In a trial of adjuvant treatment, selection of high risk patients could allow a much smaller sample size and, perhaps, identify the population most in need of treatment.
- Tumor receptor presence can predict outcome (maybe response too, a different question)

22

Selection of High Risk Patients

b. Cardiovascular

It is usual to begin outcome studies in highest risk patients, not necessarily because they benefit more (as a percent), but because you need fewer patients to obtain needed endpoints (they are also, in an older population, likely to show a larger effect on total mortality because more of their deaths will be CV)

CHF/ACEI's

- CONSENSUS (enalapril) in NYHA class III-IV patients studied only 253 patients, showing dramatic survival effect in only 6 months study. Mortality untreated was 40% in just 2 months, and treatment showed a 40% reduction.

23

Selection of High Risk Patients

b. Cardiovascular

We recognize risk stratification by LDL cholesterol, HDL cholesterol, BP, history of AMI, diabetes mellitus, and choose patients who will have higher risk, at least for initial studies.

But there are new "proteomic" measurements that seem to explain and amplify these predictors.

Heeschen, et al. JAMA 2004; 291:435-441.

Examined ability of several blood factors to predict outcome (death + AMI) in population (placebo group in CAPTURE) who all had:

- Acute Coronary Syndrome
- >70% occlusion of at least 1 coronary
- Undergone angioplasty

I.e., they all look like similar high risk patients. But they're not.

24

Selection of High Risk Patients

b. Cardiovascular

Looked at predictive value of Placental Growth Factor (PlGF)

Soluble CD40 ligand (SCD40L)

Troponin

CRP

25

Variable	HR	P
Male	0.97	.45
Diabetes	1.24	.62
Smoker	0.67	.23
Hypertension	1.03	.96
CRP	0.98	.94
Troponin>0.01 mg/L	1.83	.03
SCD40L >5mg/L	2.65	.002
PlGF >27 mg/L	3.03	<.001

First 30 days risk of Death and AMI
Cox proportional hazards
Confirmed in 600 ER chest pain patients:
PlGF>27 gave HR=4.80

26

Selection of High Risk Patients

b. Cardiovascular

These are independent risk factors so that an ACS patient with all 3 predictors would have a 14.7 fold rate of events. (Note, CRP was fully accounted for by the other measures and so were other established risk predictors: diabetes, smoking, HT, maleness.)

The potential for doing a very small study in the high risk ACS population is fairly obvious. But it also reminds us that an apparently homogenous population can have very different people in it.

PlGF is a VEGF (vascular endothelial growth factor) and may be a factor in pathological angiogenesis; SCD40L is a measure of platelet activation; and troponin indicates myocardial damage, so their predictive value is not surprising

27

Selection of High Risk Patients

c. Other

Many other possible selection mechanisms are already available or can be expected

- Genetic predictors of Alzheimer's Disease.
- Radiographic studies that suggest activity of, e.g., MS and other diseases or that predict cardiac outcomes.

28

Selection of Likely Responders

Identifying the people who will respond to a treatment, then formally studying them, greatly enhances the power of a study, facilitating approval, and also may have implications for how a drug will be used.

It can be especially critical when responders are only a small fraction of all the people with a condition, e.g., because they have the "right" receptor. In such a case finding a survival effect in an unselected population may be practically impossible.

Sometimes selection is based on understanding of the disease, i.e. pathophysiologic selection, and seems obvious.

29

Selection of Likely Responders

- Edema can result from hepatic, renal or cardiac causes. Choose the last for study of an inotrope or other cardiac intervention
- CHF can result from systolic or diastolic dysfunction. Choose the former for study of a positive inotrope, the latter for a CCB. With other kinds of drugs, e.g., diuretics or ACEIs, might stratify to see if results differ by pathophysiology

30

Selection of Likely Responders

- Hypertension can be high-renin or low-renin. High renin population would show a much larger effect than a mixed population to ACEIs, AIIBs, or BBs.
- We study antibiotics in bacterial infections sensitive to the antibacterial
- A well-established genetically determined difference could be the basis for a pathophysiologically selected population. A marker associated with a particular tumor characteristic could be a basis for selection. Most convincing so far are tumor genetics: Herceptin for Her2+ breast tumors; selection of ER^+ breast tumors for anti-estrogen treatment.

31

Selection of Likely Responders

Even if pathophysiology is unclear, likely, responders could be identified by an initial short-term response. There is a history of this:

- CAST was carried out in people who had a 70% reduction of VPB's. Only "responders" were randomized.
- Trials of topical nitrates were carried out only in people with a BP or angina response to sublingual nitroglycerin.
- Anti-arrhythmics were developed by Oates, Woosley, and Roden by open screening for response, then randomizing the responders.
- Every randomized withdrawal study has this characteristic.

32

Selection of Likely Responders

Selection could be based on response of a biomarker; that is, study the entire group and randomize only those with a good response. Possibilities

- Tumor that shows early metabolic effect on PET scan
- Tumor that shows early response on blood measure (PSA)
- Tumor that doesn't grow over an n-week period (it would be hard to randomize tumor responders to Rx vs. no Rx)
- Only patients with LDL effect > n (or some other less studied lipid)
- Only patients with CRP response > x

33

Selection of Likely Responders

We are at the very beginning of searching for genetic or other characteristics that will predict response. These could be pathophysiologic, that is, based on understanding of disease or drug mechanism (role of her 2 receptor in response to Herceptin; role of EGFR in response to erlotinib), generally with these factors identified prospectively, and with patients either selected by, or stratified by, that factor. But the selection could be simply descriptive: run a trial in unselected patients with depression, bipolar disease, lipid abnormalities, heart failure and link a genetic finding with response. In fact, would ordinarily search widely for such a relationship. Tarceva data illustrate the potential.

34

Selection of Likely Responders

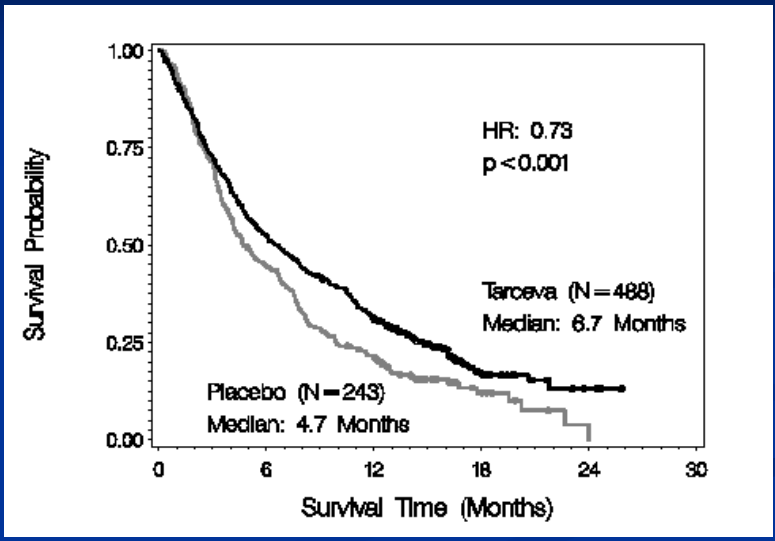
Tarceva (erlotinib)

Randomized, DB, placebo-controlled trial of Tarceva 150 mg in 731 patients with locally advanced or metastatic NSCLC after failure of ≥ 1 prior regimen. Randomized 2:1 (488 Tarceva, 243 placebo). Study overall showed clear survival effect

	Tarceva	Placebo	HR	CI
survival (mos.)	6.7	4.7	0.73	0.61-0.86 p<0.001
1 year survival	31.2%	21.5%		

35

Kaplan-Meier Curve for Overall Survival of Patients by Treatment Group



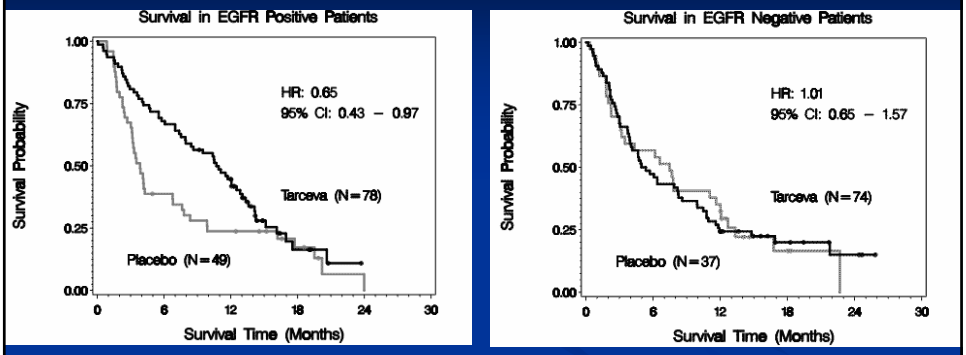
36

Tarceva (erlotinib)

Tumors were examined for EGFR expression status in 238 (of 731) patients. EGFR+ was defined as ≥10% staining using DAKO EGFR pharmDx kit.

	Tarceva	Placebo	HR	CI
EGFR+ (127)	78	49		
Survival (mos)	10.71	3.84	0.65	(0.43-0.97) p=0.033
EGFR- (111)	74	37		
Survival	5.35	7.49	1.01	(0.65-1.57) p=0.958

37



38

Practical Considerations

If a proteomic or genomic enrichment characteristic is well defined before the study, by a prior study or in some other way, there is no inferential problem. The genomically identified patients will be the only ones studied or will be a stratified group identified as the to-be-analyzed subgroup. There is still the question of how much data you need for the "off" subset.

Another critical issue is whether the genomic/proteomic identifier can be used in practice, or all patients will be treated.

But suppose, as in the Tarceva case, the subset is identified only after the study is complete.

39

Selection of Likely Responders

Such a finding would ordinarily have the properties of a retrospective subset analysis, almost never convincing by itself, but the study could be repeated with prospective stratification by the genomic marker. Or, if you were very convinced, subsequent studies of longer-term effects could be carried out in the responder population.

But Simon has proposed an alternative.

40

Simon's Proposal

1. Design study as usual, but divide into first half, second half.
2. Run first half of study and search for genetic predictor of response (any analyses, as many as you want)
3. Complete the study, entering all patients (responders predicted and not predicted) but stratifying them
4. Divide study alpha as 0.04 for whole study and 0.01 for the response-predicted subset in 2nd half.

41

Overall Strategy

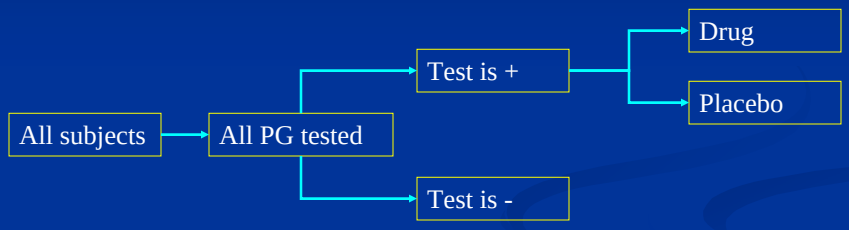
Practical Considerations

If there is a persuasive genetic/pathophysiologic marker, measurable at baseline, it seems reasonable to

- Stratify in studies by marker (+) or (-), "pre-hoc"
- Make effect in (+) the primary endpoint
- Usually, unless prior PD data make lack of effect in the (-) group completely obvious, include (-) group and evaluate effect in them as a secondary observation, looking for a difference in effect size
- If the "clear" pathophysiologic explanation arises post facto, almost always would need a confirmatory trial (or two, if explanatory feature is not completely persuasive), but a prospective plan to evaluate a positive study in two halves might be persuasive

42

Prospective, Screened - no possible effect in (-) group



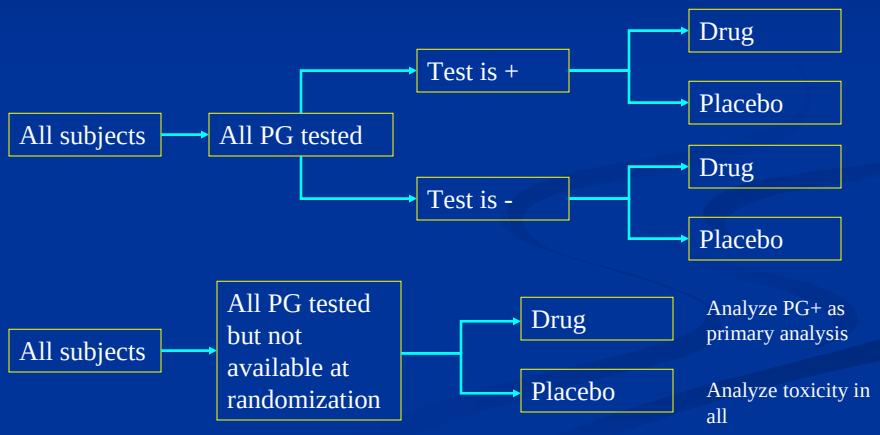
43

Considerations

- Enrichment strategy for efficacy
  - true signal of efficacy of drug - proof of principle
  - overestimate of effectiveness in an unselected population; therefore distorts B/R in that population
- Will be proof of principle and effectiveness but only if the test is available
- PG test must be available if you are not going to study the (-) group, because:
- Safety must consider all patients [(+) and (-)] if you cannot select

44

Prospective, Stratified - where there is possible effect in the (-) group and/or where toxicity in the (-) group needs to be evaluated because pre-treatment selection is not possible



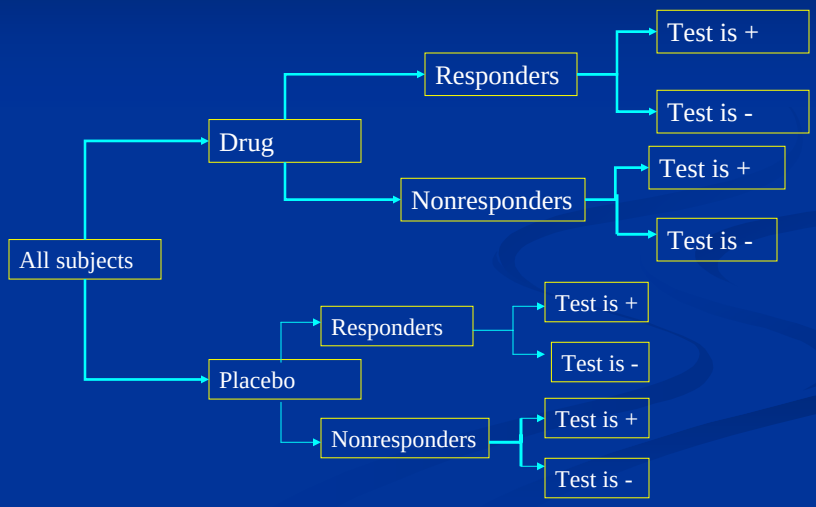
45

Considerations

- Will test efficacy in both + and - subgroups and assess safety in both
- Provides proof of principle; effect in (+) group can be the primary endpoint in both cases. If the test is not available, need to analyze the whole population for B/R and, conclude that B/R is positive for the whole population, even if only the (+) group is analyzed for effectiveness
- Needed where sensitivity cannot be assumed in advance to be very high

46

Retrospective



47

Retrospective

- Hypothesis generating; multiplicity problems
- Would usually need confirmatory clinical trial(s)
- Good basis for new stratified trial.

48

Interesting Question

Could a retrospective finding that was very consistent be persuasive without further study? For example, could 1/2 the study be used to identify genetic predictor, second 1/2 used to "confirm" it

- Seems possible, if study is positive overall and we're only talking about selection
- Effect is yes/no rather than graded, which would seem to need further evaluation
- Difference is large and highly (nominally) significant
- Simon's approach

49